Genetic loci associated with delayed clearance of *Plasmodium falciparum* following artemisinin treatment in Southeast Asia

Shannon Takala-Harrison[a], Taane G. Clark[b,1], Christopher G. Jacob[a,1], Michael P. Cummings[c,2], Olivo Miotto[d,e,2], Arjen M. Dondorp[e], Mark M. Fukuda[f], Francois Nosten[e,g], Harald Noedl[h], Mallika Imwong[i], Delia Bethell[f], Youry Se[j], Chanthap Lon[j], Stuart D. Tyner[f], David L. Saunders[f], Duong Socheat[k], Frederic Ariey[l], Aung Pyae Phyo[e,g], Peter Starzengruber[h], Hans-Peter Fuehrer[h], Paul Swoboda[h], Kasia Stepniewska[m], Jennifer Flegg[m], Cesar Arze[n], Gustavo C. Cerqueira[n], Joana C. Silva[n], Stacy M. Ricklefs[o], Stephen F. Porcella[o], Robert M. Stephens[p], Matthew Adams[a], Leo J. Kenefic[a], Susana Campino[d,q], Sarah Auburn[q], Bronwyn MacInnis[d,q], Dominic P. Kwiatkowski[d,q], Xin-zhuan Su[r], Nicholas J. White[e], Pascal Ringwald[s], and Christopher V. Plowe[a,3]

[a]Howard Hughes Medical Institute/Center for Vaccine Development, University of Maryland School of Medicine, Baltimore, MD 21201; [b]Department of Pathogen Molecular Biology, London School of Tropical Medicine, London WC1E 7HT, United Kingdom; [c]Center for Bioinformatics and Computational Biology, University of Maryland, College Park, MD 20742; [d]Medical Research Council Centre for Genomics and Global Health, University of Oxford, Oxford OX3 7BN, United Kingdom; [e]Mahidol-Oxford Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, 10400 Thailand; [f]Armed Forces Research Institute of Medical Sciences, Bangkok, 10400 Thailand; [g]Shoklo Malaria Research Unit, Mae Sod, Tak 63110, Thailand; [h]Institute of Specific Prophylaxis and Tropical Medicine, Medical University of Vienna, A-1090 Vienna, Austria; [i]Department of Molecular Tropical Medicine and Genetics, Faculty of Tropical Medicine, Mahidol University, Bangkok 10400, Thailand; [j]Armed Forces Research Institute of Medical Sciences, Phnom Penh 12252, Cambodia; [k]University of Health Science, Phnom Penh 12201, Cambodia; [l]Institut Pasteur Unité d'Immunologie Moléculaire des Parasites, F-75015 Paris, France; [m]WorldWide Antimalarial Resistance Network, Oxford University, Oxford OX3 7LJ, United Kingdom; [n]Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD 21201; [o]Research Technologies Branch, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, MT 59840; [p]Advanced Biomedical Computing Center, Frederick, MD 21702; [q]Malaria Programme, Wellcome Trust Sanger Institute, Hinxton CB10 1SA, United Kingdom; [r]Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892; and [s]Drug Resistance and Containment Unit, Global Malaria Programme, World Health Organization, 1211 Geneva 27, Switzerland

Edited* by Thomas E. Wellems, National Institutes of Health, Bethesda, MD, and approved November 7, 2012 (received for review July 6, 2012)

The recent emergence of artemisinin-resistant *Plasmodium falciparum* malaria in western Cambodia could threaten prospects for malaria elimination. Identification of the genetic basis of resistance would provide tools for molecular surveillance, aiding efforts to contain resistance. Clinical trials of artesunate efficacy were conducted in Bangladesh, in northwestern Thailand near the Myanmar border, and at two sites in western Cambodia. Parasites collected from trial participants were genotyped at 8,079 single nucleotide polymorphisms (SNPs) using a *P. falciparum*-specific SNP array. Parasite genotypes were examined for signatures of recent positive selection and association with parasite clearance phenotypes to identify regions of the genome associated with artemisinin resistance. Four SNPs on chromosomes 10 (one), 13 (two), and 14 (one) were significantly associated with delayed parasite clearance. The two SNPs on chromosome 13 are in a region of the genome that appears to be under strong recent positive selection in Cambodia. The SNPs on chromosomes 10 and 13 lie in or near genes involved in postreplication repair, a DNA damage-tolerance pathway. Replication and validation studies are needed to refine the location of loci responsible for artemisinin resistance and to understand the mechanism behind it; however, two SNPs on chromosomes 10 and 13 may be useful markers of delayed parasite clearance in surveillance for artemisinin resistance in Southeast Asia.

drug resistance | genome-wide association | molecular markers

Artemisinin-based combination therapies (ACTs) are the leading treatment for *Plasmodium falciparum* malaria (1), and their use with other tools to reduce the global malaria burden has sparked renewed consideration of malaria eradication. Malaria has been treated with artemisinin derivatives in Asia since the 1970s (2). Extremely fast-acting, artemisinins kill both young ring forms and mature blood-stage parasites (3). The World Health Organization (WHO) recommended in 2001 that artemisinins be used strictly in combination therapies in hopes of delaying the emergence of resistance (2), but ACT treatment failure rates were rising on the Thailand/Cambodia border by 2006 (4, 5) and progressively prolonged parasite clearance after treatment with artemisinin derivatives soon followed (6–8). This evidence that artemisinin resistance has emerged in western Cambodia, historically an epicenter of drug-resistant malaria, is an ominous development that threatens the recent major global investment in ACTs (3). If genetically heritable artemisinin resistance has emerged, it can be expected to follow historical patterns of antimalarial resistance (9) and disseminate globally, at immense cost to human life. Strategies for containing resistance require accurate, up-to-date information about its geographical distribution. Molecular markers of resistance would provide a practical surveillance tool.

In 2008, the WHO initiated the Artemisinin Resistance Confirmation, Characterization, and Containment (ARC3) pilot project in Southeast Asia. Four clinical trials of artesunate monotherapy were done to define clinical and parasitological responses to the artemisinins without the confounding influence of partner drugs. Trials

Author contributions: S.T.-H., A.M.D., M.M.F., F.N., H.N., M.I., D.B., Y.S., C.L., S.D.T., D.L.S., D.S., F.A., A.P.P., P. Starzengruber, H.-P.F., P. Swoboda, N.J.W., P.R., and C.V.P. designed research; S.T.-H., C.G.J., A.M.D., M.M.F., F.N., H.N., M.I., D.B., Y.S., C.L., S.D.T., D.L.S., D.S., F.A., A.P.P., P. Starzengruber, H.-P.F., P. Swoboda, S.M.R., S.F.P., M.A., L.J.K., S.C., S.A., and B.M. performed research; T.G.C., M.P.C., O.M., K.S., J.F., C.A., G.C.C., J.C.S., S.M.R., S.F.P., R.M.S., B.M., D.P.K., and X.-z.S. contributed new reagents/analytic tools; S.T.-H., T.G.C., C.G.J., M.P.C., O.M., K.S., J.F., C.A., G.C.C., J.C.S., S.M.R., S.F.P., M.A., L.J.K., S.C., S.A., and C.V.P. analyzed data; and S.T.-H., C.G.J., and C.V.P. wrote the paper.

Conflict of interest statement: As a World Health Organization (WHO) staff member, P.R. is responsible for the views expressed in this publication, which do not necessarily represent the decisions, policy, or views of the WHO. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, and mention of trade names, commercial products, or organizations does not imply endorsement by the US Government.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

Data deposition: The SNP data reported in this paper have been deposited in PlasmoDB database, http://plasmodb.org/plasmo/.

[1]T.G.C. and C.G.J. contributed equally to this work.

[2]M.P.C. and O.M. contributed equally to this work.

[3]To whom correspondence should be addressed. E-mail: cplowe@medicine.umaryland.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1211205110/-/DCSupplemental.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
01 JAN 2013	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Genetic loci associated with delayed clearance of Plasmodium falciparum following artemisinin treatment in Southeast Asia	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)	5d. PROJECT NUMBER
Takala-Harrison S., Clark T. G., Jacob C. G., Cummings M. P., Miotto O., Dondorp A. M., Fukuda M. M., Nosten F., Noedl H., Imwong M., Bethell D., Se Y., Lon C., Tyner S. D., Saunders D. L., Socheat D., Ariey F., Phyo A. P., Starzengruber P., Fuehrer H. P., Swoboda P., Stepniewska K., Flegg, J., Arze C., Cerqueira G. C., Silva J.C., Ricklefs S. M., Porcella S. F., Stephens R. M., Adams M., Kenefic L. J., Campino S., Auburn S., Macinnis B., Kwiatkowski D. P., Su X. Z., White N. J., Ringwald P., Plowe C. V.,	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
United States Army Institute ofSurgical Research, JBSA Fort Sam Houston, TX	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**6**	

were conducted at Pailin and Tasanh in western Cambodia, where emerging resistance was suspected; in Wang Pha, Thailand, on the border with Myanmar, where moderately delayed parasite clearance after artesunate-mefloquine treatment had been reported (7); and in Bandarban, Bangladesh, where artemisinins had been little used (Fig. S1). DNA extracted from *P. falciparum* parasites collected in these trials was genotyped at 8,079 single nucleotide polymorphisms (SNPs) using a *P. falciparum* SNP array. Parasite genotypes were examined for signatures of recent positive selection and association with parasite clearance phenotypes to identify regions of the parasite genome associated with artemisinin resistance.

Results

Parasite Genotypes. DNA extracted from leukocyte-depleted blood underwent whole-genome amplification and genotyping at 8,079 SNPs using a molecular inversion probe Affymetrix SNP array. Of 342 samples successfully genotyped, 331 represented clinical infections with the phenotypes of interest (Table S1). Genotypes were called from normalized signal intensities using a heuristic algorithm based on discrete cutoffs of signal strength and contrast. A total of 4,420 SNPs that were invariant or had minor allele frequency (MAF) <1% were excluded from the analysis, as were 832 SNPs with extreme numbers of undetermined or heterozygous calls. Four samples with an extreme number of undetermined SNP calls were also excluded (Fig. S2). The final analysis included 2,827 SNPs (Fig. S3), including 752 in intergenic regions, and 327 samples—200 from western Cambodia, 30 from Thailand, and 97 from Bangladesh.

Distribution and Heritability of Phenotypes. Parasite density was recorded every 6–12 h after starting artesunate treatment until infections cleared. Parasite clearance half-lives and parasite clearance times were the clinical phenotypes for genome-wide analyses. Clearance half-lives were estimated using a publicly available parasite clearance estimator (10). An ex vivo phenotype, dihydroartemisinin (DHA) IC_{50}, was also examined (Figs. S4 and S5).

Parasite clearance half-lives were markedly delayed in most patients from western Cambodia compared with Bangladesh (Fig. 1*A*). Artesunate showed a mixture of responses in Thailand, some with short half-lives comparable to those in Bangladesh and others with moderately prolonged half-lives, but none were as long as the slowest-clearing infections in Cambodia. The distribution of parasite clearance times (Fig. 1*B*) was similar to that of half-lives. These results suggest three clearance phenotypes: rapidly clearing parasites seen in all Bangladeshi and many Thai infections; moderately slow-clearing parasites seen in some Thai and a minority of Cambodian infections; and very slowly clearing parasites seen only in Cambodia, where they predominated.

ANOVA was used to assess heritability (H^2) of clearance half-life (Fig. 1*C*) and clearance time (Fig. 1*D*) in identical parasite clones observed at the two Cambodian sites (11). Different patients infected with the same clone had similar clearance half-lives and times. No Cambodian clones were seen in Thailand or Bangladesh, and no clones overlapped between Thailand and Bangladesh. Parasite clearance half-life was highly heritable (age-adjusted $H^2 = 0.52$, $F = 5.97$, $P < 0.0001$), as was clearance time (adjusted for log-transformed parasitemia at diagnosis $H^2 = 0.55$, $F = 6.54$, $P < 0.0001$), suggesting that parasite genes encoding these phenotypes may be identified by genome-wide association.

Population Structure. Population structure by geography was examined using principal components analysis (PCA; Fig. 2). The first two principal components differentiated parasites from all three countries, highlighting the need to account for population structure in tests of association. Parasites from Bangladesh clustered together; in contrast, Cambodian parasites were widely dispersed, some being more similar to Thai parasites and others clearly differentiated from Thai and from other Cambodian parasites.

Associations with Parasite Clearance Phenotypes. Using efficient mixed-model association (EMMA) (12), we used linear mixed-regression models to estimate the association between each SNP and the two clearance phenotypes (treated as continuous variables). Models included a genetic similarity matrix as a random

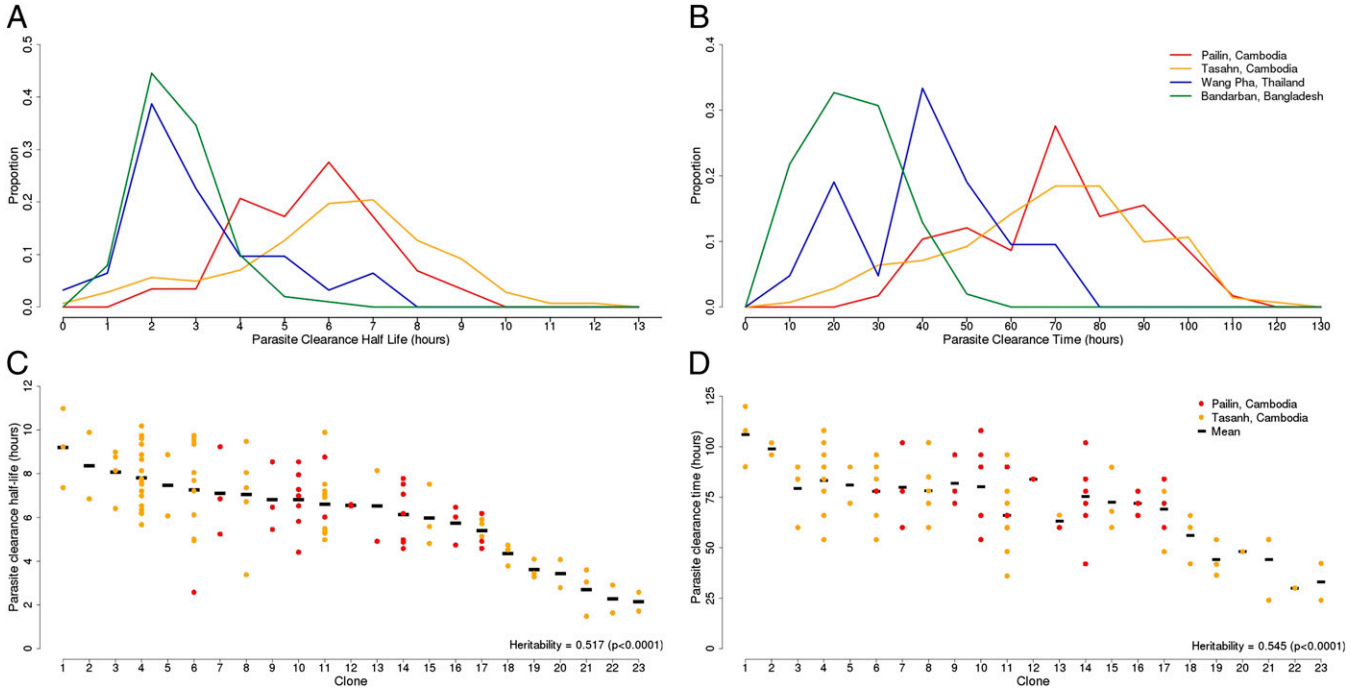

Fig. 1. Distribution and heritability of parasite clearance phenotypes. (*A* and *B*) Distribution of parasite clearance half-life and parasite clearance time at the study sites. (*C* and *D*) Distribution of clearance half-life and clearance time in 23 identical clones from Tasanh and Pailin, Cambodia. The heritability of each phenotype (adjusted for confounding factors) is shown in *C* and *D Bottom Right*.

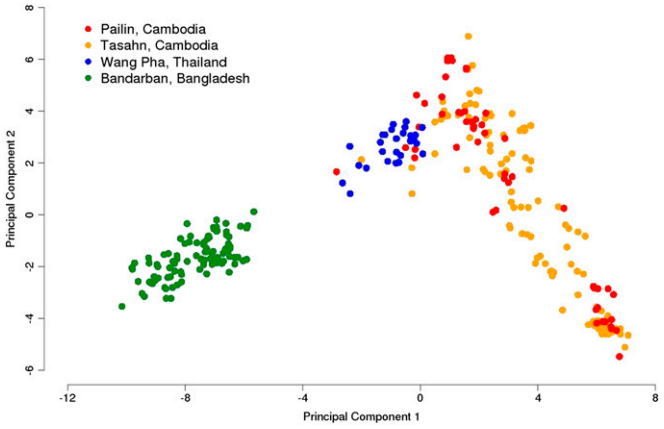

Fig. 2. Population structure by geography. A plot of the first two principal components from a principal components analysis demonstrates evidence of population structure among parasites based on geographic region.

effect to account for lack of independence among genetically similar parasites. Study site was included as a fixed effect to account for any residual confounding due to population structure. Age and parasitemia at diagnosis were also included as covariates.

Two SNPs, one on chromosome 10 (MAL10-688956) and one on chromosome 13 (MAL13-1718319), achieved genome-wide significance in models of parasite clearance half-life (Fig. 3 *A* and *B*). These SNPs and two additional ones, MAL13-1719976 and MAL14-718269, achieved genome-wide significance in models of parasite clearance time (Fig. 3 *C* and *D*). Quantile–quantile plots indicated little residual confounding due to population structure or other potential confounding variables (e.g., host immunity; Fig. 3 *B* and *D*). To further evaluate EMMA's ability to account for population structure, we repeated the same analyses excluding Bangladeshi parasites, which were clearly not resistant and genetically distinct from parasites from other sites. With Bangladeshi parasites excluded, most of the same top SNPs were observed (*SI Text*), but with increased *P* values, likely due to a loss of power from the smaller sample size or possibly to reduced variability in the phenotype after excluding a large proportion of fast-clearing parasites.

A nonparametric method, Random Forests, was also used to assess the importance of each SNP and other covariates in predicting the clinical phenotypes based on the percent increase in mean-squared error. The best predictors of clearance half-life were study site, followed by SNP MAL13-1718319 [percent variance explained by all SNPs and other covariates (%Var) = 58.4%] (Fig. S7*A*). When Bangladeshi parasites were excluded, MAL13-1718319 was the best predictor of clearance half-life (Fig. S7*B*; %Var = 38.5%). The best predictors of clearance time were study site and log-transformed parasitemia at diagnosis (Fig. S7*C*; %Var = 64.2%). The two SNPs most predictive of clearance time in the Random Forests analysis (MAL14-2492091 and MAL9-1042451) were also the most predictive of parasites, being from Bangladesh (Fig. S7*E*), suggesting these SNPs were unrelated to artemisinin resistance, rather reflecting population structure. With Bangladeshi parasites excluded, parasitemia at diagnosis remained the best predictor of clearance time, followed by SNP MAL13-1718319 (Fig. S7*D*; %Var = 30.8%; Table 1).

Markers of Delayed Parasite Clearance. To test the utility of the SNPs in Table 1 as surveillance tools for tracking parasites with delayed parasite clearance, odds ratios (ORs) were estimated at each SNP comparing the log odds of an infection having clearance half-life >5 h (the median clearance half-life in our dataset) in parasites with a given allele to those with the alternative allele (Table 2). The proportion of infections with half-life >5 h was 77% in western Cambodia, 20% in Thailand, and 3.1% in Bangladesh. The frequency of the A and T alleles at SNPs MAL10-688956 and MAL13-1718319, respectively, closely mirrored the frequency of delayed clearance at each site (Table 2). The odds of delayed clearance were also significantly greater in infections with these alleles compared with those with the alternative allele in the full dataset [OR = 15.1 and OR = 22.8 for MAL10-688956 (A) and MAL13-1718319 (T), respectively] and in Cambodian parasites alone [OR = 6.1 and OR = 6.7 for MAL10-688956

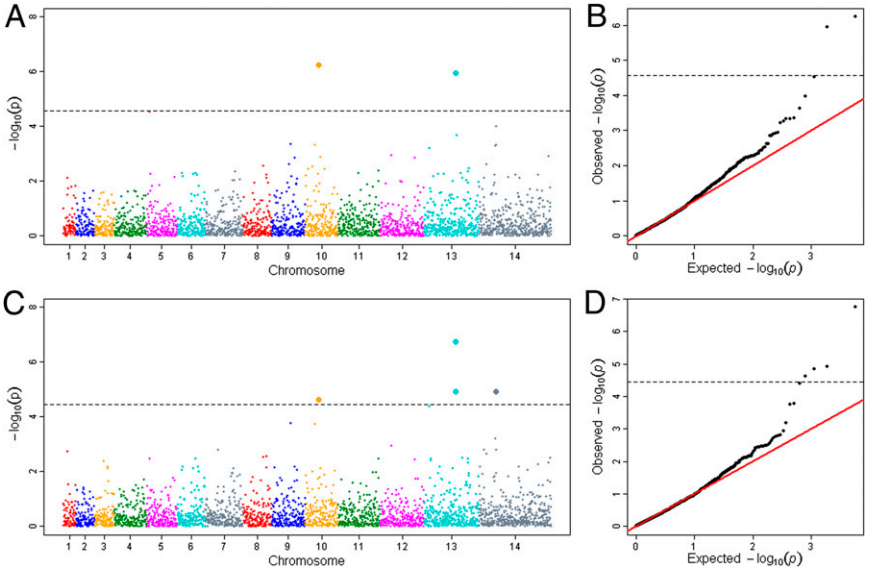

Fig. 3. SNPs associated with parasite clearance in regression models. Manhattan plots showing $-\log_{10}$ *P* values for each SNP tested in EMMA models of (*A*) parasite clearance half-life and (*C*) parasite clearance time for the dataset including all study samples. Corresponding quantile–quantile plots for models of (*B*) parasite clearance half-life and (*D*) clearance time are also shown. The red lines in *B* and *D* indicate the distribution of expected *P* values in the absence of any association or confounding. Early deviation from the expected line suggests the presence of confounding that has not been adequately controlled for in the analysis, whereas deviations at the highest $-\log_{10}$ *P*-value ranges represent the loci most strongly associated with the phenotype. In all panels, the dashed horizontal lines indicate thresholds for genome-wide significance determined using a phenotype-permutation approach.

Table 1. SNPs associated with parasite clearance phenotypes

		Clearance half-life		Clearance time	
Chromosome	SNP	Random Forests rank	*P* value	Random Forests rank	*P* value
10	MAL10-688956	11	**5.8E-07**	>20	**2.42E-05**
13	MAL13-1718319*	1	**1.1E-06**	1	**1.81E-07**
	MAL13-1719976*	5	2.3E-04	2	**1.41E-05**
14	MAL14-718269	>20	1.1E-04	>20	**1.18E-05**

Bold font indicates SNPs that reached genome-wide significance in the EMMA models including all study samples. Random Forests rank represents the rank of genetic marker predictors only, with parasites from Bangladesh excluded to reduce confounding due to population structure.
*SNPs located within a top-ranked signature of recent positive selection.

(A) and MAL13-1718319 (T), respectively] (Table 2). MAL13-1719976 (C) had a high prevalence in all three countries, including Bangladesh, where parasites with delayed clearance were infrequent. MAL14-718269 (T) showed weaker associations with delayed clearance compared with the SNPs on chromosomes 10 and 13 (Table 2).

Signatures of Recent Positive Selection. To locate regions of the parasite genome under recent positive selection in Cambodian parasites, cross-population extended haplotype homozygosity (XP-EHH) was used to identify high-frequency alleles associated with long haplotypes (13), and pairwise F_{ST} to identify loci that diverged significantly between sites (14). Parasites from the two western Cambodia sites were grouped together based on their sharing of genetically identical parasite clones (Fig. 1). Thailand and Bangladesh served as comparator populations. Based on combined F_{ST}–XP-EHH scores (Fig. S8), multiple regions showed positive selection in Cambodian parasites (Fig. S9), including regions containing genes associated with resistance to other antimalarial drugs (*pfcrt*, *pfmdr1*, *dhps*, and *dhfr*). The strongest signatures (representing the top 10% of SNP windows from each site comparison) were located on chromosomes 4, 5, 6, 7, 9, 10, 13, and 14 (Fig. S9 and Table S2). One of these top-ranked regions, a 239-kb region of chromosome 13, contained SNPs associated with artemisinin resistance phenotypes (Table 1) MAL13-1718319 and MAL13-1719976 (Fig. S9). SNPs MAL10-688956 and MAL14-718269 (Table 1) were also in regions that appeared to be under recent positive selection, albeit not in the top-ranked regions (Figs. S9 and S10).

Genes in Regions Associated with Artemisinin Resistance. Two SNPs associated with parasite clearance phenotypes are in intergenic regions and two are nonsynonymous SNPs within genes. MAL13-1718319 is in a gene encoding a RAD5 homolog (MAL13P1.216), and MAL14-718269 is in a pseudogene encoding a cyclic nucleotide-binding protein (PF14_0173). MAL10-688956 is in the 3′ untranslated region of the gene encoding DNA polymerase delta catalytic subunit (PF10_0165), whereas MAL13-1719976 is 690 bp downstream from the gene encoding the RAD5 homolog containing SNP MAL13-1718319.

To define linkage disequilibrium (LD) windows around top genome-wide association study (GWAS) "hits" in which to seek candidate genes, we used Haploview (15). The three LD windows containing SNPs associated with the clinical phenotypes are shown in Table S3; both chromosome 13 SNPs are in the same window. Gene ontology functions for 53 genes in the three LD windows containing SNPs associated with clinical phenotypes are listed in Dataset S1; 21 of these genes were located in regions of chromosomes 13 and 14, representing the top-ranked signatures of selection (Dataset S2).

Discussion

In this initial GWAS of *P. falciparum* to examine associations with delayed parasite clearance, both parasite clearance half-life and clearance time following artesunate treatment were found to be heritable, suggesting that genes underlying these phenotypes may be identified by genome-wide association. Four SNPs, on chromosomes 10, 13 (two SNPs), and 14 were significantly associated with parasite clearance phenotypes. The two SNPs on chromosome 13 are in a region of the genome that appears to have been under strong recent positive selection in Cambodia. Though replication and validation studies are needed to identify precisely the loci responsible for artemisinin resistance and the mechanism(s) behind it, two SNPs identified in this study (on chromosomes 10 and 13) may prove useful as markers of delayed parasite clearance in Southeast Asia.

Our analyses showed clear population structure by geography, but unlike parasites from Thailand and Bangladesh, Cambodian parasites did not cluster tightly in the PCA. The complex parasite population structure observed in Cambodia cannot be explained by grouping parasites from the two Cambodian sites, which are in close geographic proximity, and the same structure patterns were observed within each site individually. Previous GWAS examining in vitro susceptibility to artemisinins have also noted multiple populations of parasites in Cambodia (16, 17). This complex population structure in Cambodia warrants further investigation.

Table 2. Association of *P. falciparum* SNPs with delayed parasite clearance after treatment with artesunate

		Allele frequency, %*			OR (95% confidence interval)†	
Chromosome	SNP (allele)	Cambodia	Thailand	Bangladesh	Cambodia only	All sites
10	MAL10-688956 (A)	0.61	0.29	0.01	6.1 (2.9–12.6)	15.1 (8.4–27.1)
13	MAL13-1718319 (T)	0.70	0.14	0.00	6.7 (3.3–13.6)	22.8 (12.6–41.2)
	MAL13-1719976 (C)	0.86	0.62	0.97	7.5 (3.2–17.7)	3.2 (1.6–6.6)
14	MAL14-718269 (T)	0.30	0.00	0.00	1.0 (0.5–2.2)	4.6 (2.4–9.0)

*Proportion of infections from Cambodia (n = 200), Thailand (n = 30), and Bangladesh (n = 97).
†Log odds of an infection having half-life >5 h in infections with the allele in noted in parentheses compared with infections with the alternative allele. Proportion of infections with clearance half-life > 5 h: 77% (Cambodia), 20% (Thailand), and 3.1% (Bangladesh).

To identify genomic regions in which to seek candidate artemisinin resistance genes, we defined LD windows around SNPs associated with parasite clearance in Cambodia. These LD windows were wide (76–94 kb), containing large numbers of genes, consistent with extended haplotypes seen with recent selection. No significant SNPs showed $r^2 > 0.5$ with any adjacent SNPs with MAF >0.05. Whole-genome sequencing of 227 *P. falciparum* field isolates indicated that r^2 fell below 0.10 within 1 kb in all parasite populations studied (18). The markers in our GWAS were, on average, 7 kb apart, so it may be that other loci associated with parasite clearance were not detected because they were not in LD with a genotyped SNP. This possibility and the relatively small sample size of this initial study (compared with typical human GWAS) highlight the need for larger replication studies with denser SNP coverage to map artemisinin resistance loci more finely.

Earlier GWAS using in vitro resistance phenotypes (16, 17, 19) found associations between SNPs in several genes and IC_{50} of DHA and other artemisinin derivatives, but none were in our top-ranked signatures of selection or in LD windows corresponding to our GWAS hits. This discordance is likely explained by the use of different resistance phenotypes. Previous studies assessed an in vitro phenotype in culture-adapted parasites continuously exposed to drug, whereas we studied clinical phenotypes in patients with acute malaria treated with rapidly cleared daily artesunate. In our study, DHA IC_{50} had low heritability, was significantly associated with no SNPs, and did not overlap with candidate SNPs associated with clinical parasite clearance phenotypes. These findings confirm that DHA IC_{50} does not capture the delayed clearance phenotype that is the hallmark of artemisinin resistance in Cambodia (6–8) and emphasize the need to develop in vitro assays of artemisinin susceptibility that can be used for functional validation of candidate resistance genes.

Agreement was seen, however, between regions under recent positive selection in our study and those identified recently by others as being under strong recent selection in Cambodia (20). Cheeseman et al. (20) estimated F_{ST} and XP-EHH for 6,969 genome-wide SNPs and identified 33 genomic regions under recent selection in Cambodia compared with Thailand and Laos, 11 of which overlapped with top-ranked signatures of selection in our study. Of 90 SNPs genotyped by Cheeseman et al. (20) within the selected regions, two on chromosome 13 were associated with parasite clearance rates in northwestern Thailand. Fine-mapping of a 550-kb region surrounding these SNPs identified a 35-kb region ~45 kb downstream from MAL13-1718319 and MAL13-1719976, SNPs associated with delayed parasite clearance in our study. Although the LD window containing these SNPs does not overlap with the region identified by Cheeseman et al. (20), both regions are within the top-ranked signature of selection on chromosome 13 identified in our study. Because the chromosome 13 signature identified in the Cheeseman et al. study (20) began at position 1735000, and associations were examined in selected regions only, association with SNPs MAL13-1718319 and MAL13-1719976 could not have been identified in that study. In contrast, our study included five SNPs within the locus identified by Cheeseman et al. (20), none of which were associated with parasite clearance. Further investigation is needed to confirm the presence of a causal locus on chromosome 13 and refine its location.

Interestingly, three of the four SNPs associated with parasite clearance in our study lie in or near genes involved in the same metabolic pathway. MAL10-688956 is located in the 3′ untranslated region of DNA polymerase delta (PF10_0165), and MAL13-1718319 and MAL13-1719976 are located in or near a RAD5 homolog (MAL13P1.216). Both of these proteins are thought to be involved in postreplication repair (PRR) (21), a DNA-damage tolerance pathway. RAD5 is thought to polyubiquitinate proliferating cell nuclear antigen (a DNA clamp that assists in processivity of DNA polymerase delta) in conjunction with two ubiquitin-conjugating enzymes, prompting activation of error-free DNA repair via template switching (21). Also involved in this pathway are several deubiquitinating enzymes, including *ubp1*, which has been linked to artemisinin resistance in the rodent malaria *Plasmodium chabaudi* (22). This pathway may be activated by DNA damage caused by oxidative stress from toxic by-products of hemoglobin degradation following artemisinin treatment (23). Mutations in this pathway could result in cell cycle arrest, as was observed in a PRR knockout model in yeast (24). Such down-regulation of the cell cycle is consistent with reduced DNA synthesis and other metabolic functions observed in the ring and trophozoite stages of parasites showing delayed clearance following artemisinin treatment (25).

The gene ontology for PF13_0237, immediately upstream from the RAD5 homolog on chromosome 13, also suggests a possible function in DNA replication and cell cycle regulation. This functional assignment is based on a protein domain similar to Cdt1, a replication initiation factor essential for cell cycle progression (26). The role of these proteins and pathways in artemisinin resistance is plausible, but needs further evaluation.

Resistance-associated alleles at SNPs MAL13-1718319 and MAL10-688956 had frequencies closely mirroring the prevalence of parasites with delayed clearance at the study sites, supporting the notion that they could be useful markers for surveillance of artemisinin resistance. However, these SNPs are evidently not sufficient to cause resistance, because they are found in other *P. falciparum* strains [e.g., V1/S and IT (MAL13-1718319-T) and V1/S, IT, 106/1, and FCR3 (MAL10-688956-A)] of diverse geographic origins collected well before reports of artemisinin resistance in Asia. These SNPs may be in LD with causal loci that were not on the SNP array; their utility as surveillance tools requires validation in studies done in diverse geographic regions with a range of parasite clearance half-lives. Rapid molecular assays suitable for typing these two candidate markers using dried blood spots are available at www.wwarn.org/toolkit/procedures/molecular.

SNPs called from whole-genome sequencing of parasites (18) collected from ARC3 and subsequent artesunate efficacy trials will provide data for larger replication studies with denser SNP coverage. The genomic and clinical data and cryopreserved parasites from these studies will also facilitate candidate gene association studies and functional validation of candidate genes.

Materials and Methods

Please see *SI Text* for a more detailed description of the methods used in this study.

ARC3 Clinical Trials. Clinical trials of artesunate efficacy were conducted at two sites in western Cambodia (Pailin and Tasanh), and one site each in northwestern Thailand (Wang Pha) and Bangladesh (Bandarban) following protocols approved by the Research Ethics Review Committee of the World Health Organization, as well as local Institutional Review Boards at each study site. Details of each trial are shown in Table S1.

Parasite Genotyping. DNA was extracted from leukocyte-depleted blood and underwent whole-genome amplification (WGA) and genotyped at 8,079 SNPs using a molecular inversion probe Affymetrix *P. falciparum* SNP array. The assay was performed following the Affymetrix GeneChip Scanner 3000 User Guide without modification, except for alterations made to the first PCR thermal cycling parameters (16).

Phenotypes. ***Parasite clearance time and rate.*** Parasite clearance half-lives were estimated using a parasite clearance estimator developed by the WorldWide Antimalarial Resistance Network (10). The estimator calculates parasite clearance rate based on the linear portion of the $\log_e$ parasitemia-time curve, and half-life is estimated as $\log_e(2)$/clearance rate. Please see *SI Text* for details of susceptibility testing.

Data Analysis. ***Genotype calling and quality control.*** Raw allele intensity data from the SNP array were normalized using the Affymetrix GeneChip Targeted Genotyping Analysis Software (GTGS). To assess robustness of SNP calls, genotypes were called using three algorithms: (*i*) GTGS, (*ii*) illuminus (27), and (*iii*) a heuristic algorithm based on discrete cutoffs of intensity strength and contrast, with cutoffs established by analysis of empirical distributions. SNP calls from the three algorithms were >90% concordant, and calls from the heuristic algorithm were used in the analysis. Genotype data has been submitted to PlasmoDB for public access (http://plasmodb.org/plasmo/). SNPs

that were invariant or had MAF <1% were excluded from the analysis, as were SNPs with an extreme number of undetermined or heterozygous calls (>15%). Samples with an extreme number of undetermined SNP calls (>20%) were also excluded. Thresholds used to select SNPs and samples for analysis are detailed in Fig. S2.

Phenotype heritability. ANOVA was used to assess heritability (H^2) of phenotypes in identical parasite clones infecting multiple individuals (11). Clones were identical at all of the 8,079 SNPs on the array except heterozygous SNPs and those missing in pairwise comparisons.

PCA and regression. Population structure was evaluated by PCA (28). We used linear mixed-regression models to estimate the effect of each SNP on parasite clearance half-life and clearance time, both of which were treated as continuous variables. Age, study site, and log-transformed parasitemia at diagnosis were included as fixed effects, and a genetic similarity matrix was included as a random effect to account for population structure. EMMA was used to determine restricted maximum-likelihood estimates and *P* values (12). Quantile–quantile plots for *P* values were used to assess the robustness of modeling approaches at minimizing false positive results due to population structure, and indicated that EMMA accounted for population structure more effectively than did PCA (Fig. S6). Because *Plasmodium* is haploid, heterozygous calls were excluded when estimating genotype/phenotype associations. Phenotype permutation was used to estimate the threshold for genome-wide significance, yielding thresholds for clearance half-life and clearance time of 2.7E-05 and 3.6E-05, respectively. The quality of genotype calling for significant SNPs was assessed by inspection of cluster plots (29). Regression analysis and PCA were performed using R statistical software (30).

Random Forests. Random Forests analyses (31) were done using the randomForest package (32) of R. A total of 1,031 variables were tried at each split, with 10,000 trees in the forest. The importance of each SNP in predicting the clearance phenotypes (treated as continuous variables) was assessed based on the percent increase in mean-squared error.

Odds ratios to evaluate SNPs as markers of delayed parasite clearance. Logistic regression was used to estimate odds ratios comparing the log odds of an infection having clearance half-life >5 h (the median clearance half-life in our data set) in parasites with a given allele at each SNP compared with infections with the alternative allele.

Definition of LD windows around phenotype-associated SNPs. LD windows containing SNPs associated with clinical phenotypes were defined using Haploview (15). LD windows were defined to include all SNPs upstream and downstream from the associated SNP up to but not including the next SNP with MAF >0.05 and R^2 <0.3. LD windows were defined in parasites in each country separately to avoid false estimates of LD due to population structure.

Signatures of selection. Parasites were grouped into populations based on country. SNPs and samples with a large proportion of missing data or heterozygous calls were excluded. No MAF threshold was imposed. A total of 6,649 SNPs were used in the analysis. XP-EHH was calculated (13). Core SNPs included all SNPs with MAF >0.05, with haplotypes examined 500 kb in the forward and reverse directions. All XP-EHH scores were binned and normalized. Wright's F_{ST} statistic was calculated for each SNP in pairwise population comparisons (14). SNP F_{ST} values were averaged within 20-SNP sliding windows, with a five-SNP overlap (33, 34). Genomic regions under recent selection in Cambodia were identified based on combined XP-EHH–F_{ST} scores, calculated within the same SNP windows used in the sliding window F_{ST} analysis. XP-EHH–F_{ST} scores were determined independently for Cambodia/Thailand and Cambodia/Bangladesh comparisons. Windows representing regions under selection in Cambodia (Fig. S10) were selected based on the inflection point of a plot of the ranked XP-EHH–F_{ST} scores from both site comparisons (Fig. S9), and top-ranked signatures were selected from among these regions as the top 10% of windows from each site comparison.

ACKNOWLEDGMENTS. This work was supported by the Bill and Melinda Gates Foundation; the Howard Hughes Medical Institute; the Doris Duke Charitable Foundation; the Division of Intramural Research, National Institute for Allergy and Infectious Diseases; National Institutes of Health (NIH) Grant IH K12RR023250 (to S.T.-H.); National Cancer Institute and NIH Contract HHSN261200800001E (to R.M.S.); and Centre for Genomics and Global Health by the Medical Research Council Grants G19/9, G0600230, and G0600718 (to D.P.K. and O.M.). The Wellcome Trust provides core support to the Wellcome Trust Sanger Institute under Contract 077012/Z/05/Z, 098051, and the Resource Centre for Genomic Epidemiology of Malaria Contract 090770/Z/09/Z.

1. World Health Organization (2011) *World Malaria Report 2011* (WHO, Geneva).
2. World Health Organization (2001) *Antimalarial Drug Combination Therapy. Report of a WHO Technical Consultation* (WHO, Geneva).
3. Dondorp AM, et al. (2011) The threat of artemisinin-resistant malaria. *N Engl J Med* 365(12):1073–1075.
4. Alker AP, et al. (2007) Pfmdr1 and in vivo resistance to artesunate-mefloquine in falciparum malaria on the Cambodian-Thai border. *Am J Trop Med Hyg* 76(4): 641–647.
5. Denis MB, et al. (2006) Surveillance of the efficacy of artesunate and mefloquine combination for the treatment of uncomplicated falciparum malaria in Cambodia. *Trop Med Int Health* 11(9):1360–1366.
6. Bethell D, et al. (2011) Artesunate dose escalation for the treatment of uncomplicated malaria in a region of reported artemisinin resistance: A randomized clinical trial. *PLoS ONE* 6(5):e19283.
7. Dondorp AM, et al. (2009) Artemisinin resistance in *Plasmodium falciparum* malaria. *N Engl J Med* 361(5):455–467.
8. Noedl H, et al.; Artemisinin Resistance in Cambodia 1 (ARC1) Study Consortium (2008) Evidence of artemisinin-resistant malaria in western Cambodia. *N Engl J Med* 359(24): 2619–2620.
9. Plowe CV (2009) The evolution of drug-resistant malaria. *Trans R Soc Trop Med Hyg* 103(Suppl 1):S11–S14.
10. Flegg JA, Guerin PJ, White NJ, Stepniewska K (2011) Standardizing the measurement of parasite clearance in falciparum malaria: The parasite clearance estimator. *Malar J* 10:339.
11. Anderson TJ, et al. (2010) High heritability of malaria parasite clearance rate indicates a genetic basis for artemisinin resistance in western Cambodia. *J Infect Dis* 201(9): 1326–1330.
12. Kang HM, et al. (2008) Efficient control of population structure in model organism association mapping. *Genetics* 178(3):1709–1723.
13. Sabeti PC, et al.; International HapMap Consortium (2007) Genome-wide detection and characterization of positive selection in human populations. *Nature* 449(7164): 913–918.
14. Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* 74(1):175–195.
15. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* 21(2):263–265.
16. Mu J, et al. (2010) *Plasmodium falciparum* genome-wide scans for positive selection, recombination hot spots and resistance to antimalarial drugs. *Nat Genet* 42(3): 268–271.
17. Yuan J, et al. (2011) Chemical genomic profiling for antimalarial therapies, response signatures, and molecular targets. *Science* 333(6043):724–729.
18. Manske M, et al. (2012) Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature* 487(7407):375–379.
19. Van Tyne D, et al. (2011) Identification and functional validation of the novel antimalarial resistance locus PF10_0355 in *Plasmodium falciparum*. *PLoS Genet* 7(4): e1001383.
20. Cheeseman IH, et al. (2012) A major genome region underlying artemisinin resistance in malaria. *Science* 336(6077):79–82.
21. Ginsburg H (2006) Progress in in silico functional genomics: The malaria Metabolic Pathways database. *Trends Parasitol* 22(6):238–240.
22. Hunt P, et al. (2007) Gene encoding a deubiquitinating enzyme is mutated in artesunate- and chloroquine-resistant rodent malaria parasites. *Mol Microbiol* 65(1): 27–40.
23. Klonis N, et al. (2011) Artemisinin activity against *Plasmodium falciparum* requires hemoglobin uptake and digestion. *Proc Natl Acad Sci USA* 108(28):11405–11410.
24. Hishida T, Kubota Y, Carr AM, Iwasaki H (2009) RAD6-RAD18-RAD5-pathway-dependent tolerance to chronic low-dose ultraviolet light. *Nature* 457(7229):612–615.
25. Mok S, et al. (2011) Artemisinin resistance in *Plasmodium falciparum* is associated with an altered temporal pattern of transcription. *BMC Genomics* 12:391.
26. Xouri G, Dimaki M, Bastiaens PI, Lygerou Z (2007) Cdt1 interactions in the licensing process: A model for dynamic spatiotemporal control of licensing. *Cell Cycle* 6(13): 1549–1552.
27. Teo YY, et al. (2007) A genotype calling algorithm for the Illumina BeadArray platform. *Bioinformatics* 23(20):2741–2746.
28. Price AL, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38(8):904–909.
29. Jallow M, et al.; Wellcome Trust Case Control Consortium; Malaria Genomic Epidemiology Network (2009) Genome-wide and fine-resolution association analysis of malaria in West Africa. *Nat Genet* 41(6):657–665.
30. R Development Core Team (2008) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna).
31. Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32.
32. Liaw A, Wiener M (2002) Classification and regression by randomForest. *R News* 2(3): 18–22.
33. Oleksyk TK, et al. (2008) Identifying selected regions from heterozygosity and divergence using a light-coverage genomic dataset from two human populations. *PLoS ONE* 3:, e1712.
34. Weir BS, Cardon LR, Anderson AD, Nielsen DM, Hill WG (2005) Measures of human population structure show heterogeneity among genomic regions. *Genome Res* 15(11):1468–1476.